Molecular Variability of the 5′- and 3′-Terminal Regions of Citrus Tristeza Virus RNA

C. López, M. A. Ayllón, J. Navas-Castillo, J. Guerri, P. Moreno, and R. Flores

First and sixth authors: Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Universidad Politécnica de Valencia, Camino de Vera 14, 46022 Valencia, Spain; and second, third, fourth, and fifth authors: Instituto Valenciano de Investigaciones Agrarias (IVIA), Apdo Oficial, 46113 Moncada, Valencia, Spain.

Accepted for publication 30 March 1998.

ABSTRACT

López, C., Ayllón, M. A., Navas-Castillo, J., Guerri, J., Moreno, P., and Flores, R. 1998. Molecular variability of the 5′- and 3′-terminal regions of citrus tristeza virus RNA. Phytopathology 88:685-691.

Isolates of citrus tristeza virus (CTV) differ widely in their biological properties. These properties may depend on the structure of viral RNA populations comprising the different isolates. As a first approach to study the molecular basis of the biological variability, we have compared the sequences of multiple cDNA clones of the two terminal regions of the RNA from different CTV isolates. The polymorphism of the 5′ untranslated region (UTR) allowed the classification of the sequences into three groups, with intragroup sequence identity higher than 88% and intergroup sequence identity as low as 44%. The variability of an open reading frame (ORF) 1a segment adjacent to the 5′ UTR supports the same grouping. Some CTV isolates contained sequences of more than one group. Most sequences from Spanish isolates belonged to group III, whereas a Japanese isolate was composed mostly of sequences of groups I and II. The mildest isolates contained only sequences of group III, whereas the most severe isolates also contained sequences of groups I, II, or both. The most stable secondary structure predicted for the 5′ UTR was composed of two stem-loops and remained essentially unchanged as a result of compensatory mutations in the stems and accommodation of most of the variability in the loops. In contrast to the 5′-terminal region, the variability of the 3′-terminal region of CTV RNA was very much restricted, with nucleotide identity values higher than 90%. The presence of a conserved putative "zinc-finger" domain adjacent to a basic region in p23, the predicted product of ORF 11, suggests that this protein might act as a regulatory factor during virus replication.

Citrus tristeza virus (CTV), the causal agent of one of the most economically important diseases of citrus, has a positive-sense RNA genome encapsidated in flexuous particles about 2,000 nm long (4). Virions contain two capsid proteins, a 25-kDa coat protein (26,29) covering about 95% of the particle length and a diverged 27-kDa coat protein (11,26) that covers one end of the particle, forming a "rattlesnake" structure (1).

Two complete genomic sequences from CTV isolates T36 from Florida (14,26) and VT from Israel (19) have been reported recently and found to contain 19,296 and 19,226 nucleotides (nt), respectively, organized in 12 open reading frames (ORFs) potentially encoding at least 17 protein products. Surprisingly, comparisons between T36 and VT sequences showed an asymmetric distribution of nucleotide differences along the genome. While the 3′ half of the genomic RNA was relatively conserved (approximately 90% identity), less than 70% identity was observed in the 5′ end, an unexpected sequence relationship between two strains of the same virus (18,19). In contrast with the wide nucleotide differences detected in the 5′ end of the T36 and VT reported sequences, their pathogenic behavior in various citrus species is not too different. Thus, both isolates cause decline of sweet orange (*Citrus sinensis* (L.) Osb.) grafted on sour orange (*C. aurantium* L.) rootstock, moderate to severe symptoms on Mexican lime (*C. aurantifolia* (Christm.) Swing.), and moderate to mild seedling yellows on sour orange and Eureka lemon (*C. limon* (L.) Burm. f.) seedlings (5,28).

Corresponding author: R. Flores; E-mail address: rflores@ibmcp.upv.es

EMBL accession numbers for the 5′ untranslated region and first 159 nucleotides of eight clones from five citrus tristeza virus isolates are from Y16838 to Y16845.

Publication no. P-1998-0507-02R

The two reported CTV sequences are consensus sequences derived from a number of cDNA clones. However, as with other RNA viruses, CTV isolates are expected to contain a spectrum of related sequence variants as a consequence of the error-prone nature of the enzymes catalyzing viral RNA replication (9). The population complexity of isolates can be further increased in aphid-transmissible viruses such as CTV due to successive infections of the same plant with different inocula. Biological properties may be influenced by the structure of viral RNA populations and, therefore, identifying the nature of the sequence variants present in them is a first step essential to understanding important features such as pathogenicity.

Because the population structure of CTV isolates has not been explored, we have compared the sequence variants from a series of CTV isolates differing in symptom expression and geographical origin. Due to the large size of CTV RNA, a detailed analysis of the complete genome was not feasible, and therefore, we have restricted our study to 1,250 nt representing approximately 6.5% of the genomic RNA and encompassing its two terminal regions. These nucleotides include the 5′ untranslated region (UTR) of 107 nt and the adjacent first 159 nt of ORF 1a, as well as the 3′ UTR of 273 nt, the flanking ORF 11 (p23) of 630 nt, and the upstream intercistronic region of 81 nt (14,26). To evaluate the genetic diversity of CTV RNA, we have sequenced several clones within each isolate. An interpretation of the unusual sequence divergence found at the 5′ end of the RNA of CTV isolates is also discussed.

MATERIALS AND METHODS

Virus isolates. CTV isolates T308, T388, T385, T317, T318, T411, T408, T373, and T346 belong to a collection of citrus virus isolates kept at the Instituto Valenciano de Investigaciones Agrarias in Moncada, Valencia. T308 comes from a calamondin (*C. madurensis* Lour.) tree of unrecorded origin in the citrus variety collection of the former Estación Naranjera in Burjassot, Valencia. It

induces intense vein-clearing and stem-pitting in Mexican lime, but is asymptomatic in sweet orange and does not cause seedling yellows (2). T388 was found in an early Satsuma mandarin (*C. unshiu* (Mack.) Marc.) illegally imported from Japan and induces seedling yellows on grapefruit (*C. paradisi* Macf.) and sour orange and severe stem-pitting in various citrus species including Mexican lime and sweet orange seedlings (3). T385 was obtained from a symptomless old tree in Orihuela (Southern Alicante, Spain); it only induces mild vein-clearing in Mexican lime (21). T317 and T411 were subcultured from T385 by aphid- and graft-inoculation, respectively, to Mexican lime and then graft-transmitted to Etrog citron (*C. medica* L.) plants. T318 was obtained by graft-transmission of T317 to Pineapple sweet orange. Symptoms induced by T317 and T411 resembled those of T385, whereas T318 caused severe symptoms similar to those of T388 (21,22). T408, T373, and T346 were obtained from Southern Valencia, Southern Murcia, and Sevilla (Spain), respectively, and cause moderate symptoms in Mexican lime and decline of sweet orange grafted on sour orange, but are symptomless on sweet orange or grapefruit seedlings. A summary of the biological properties of these isolates is presented in Table 1.

Isolation of double-stranded RNA (dsRNA) from infected tissue. Bark from vigorous sweet orange or Etrog citron young shoots was pulverized with liquid nitrogen. Nucleic acid preparations enriched in dsRNAs were obtained by extraction with buffer-saturated phenol and fractionation by column chromatography on nonionic cellulose (CF-11; Whatman International, Maidstone, England) as previously described (23).

cDNA synthesis and polymerase chain reaction (PCR) amplification. First-strand cDNA was synthesized using as templates total CTV-specific dsRNAs or full-genome-length CTV dsRNA eluted from low-melting-point agarose gels. The dsRNAs were denatured with 10 mM methylmercuric hydroxide and polyadenylated using yeast poly(A) polymerase (United States Biochemical Corp., Cleveland) according to the manufacturer's instructions. After phenol-chloroform extraction and ethanol precipitation, the polyadenylated RNAs were reverse-transcribed by incubation at 42°C for 60 min in a reaction mixture (20 µl) containing 1× AMV buffer (50 mM Tris-HCl, pH 8.5; 30 mM KCl; 8 mM $MgCl_2$; and 1 mM DTT), 1 mM each of the four dNTPs, 10 U of RNasin (Amersham Life Science, Amersham, United Kingdom), 25 U of avian myeloblastosis virus reverse transcriptase (AMV-RT) (Boehringer GmbH, Mannheim, Germany), and 100 ng of primer PM-1 with a dT_{17} at its 3′ end (Table 2). For synthesis of second-strand cDNA, an aliquot ($^1/_{20}$) of this preparation was PCR-amplified in a reaction mixture (50 µl) containing 1× PCR buffer (10 mM Tris-HCl, pH 8.3; 50 mM KCl; and 1.5 mM $MgCl_2$,), 200 µM each of the four dNTPs; 2.5 U of *Taq* DNA polymerase (Boehringer GmbH), and 500 ng each of primers PM-1 and RF-92, to amplify the 5′ end of the viral RNA, and PM-1 and RF-72, to amplify the 3′ end of the viral RNA (Table 2). Primers RF-92 and RF-72 were designed selecting regions of the CTV RNA essentially conserved in T36 (14,26) and VT (19) and in several Spanish isolates as revealed by previous RT-PCR amplifications (data not shown). The PCR cycling profile (30 cycles) was 94°C for 45 s, 55°C for 45 s, and 72°C for 2 min, with a final extension step at 72°C for 10 min. PCR-amplified products were separated by agarose or polyacrylamide gel electrophoresis.

Cloning and sequencing. DNAs of the expected size were eluted and directly cloned in the linearized and thymidylated pT7Blue(R) plasmid (Novagen, Inc., Madison, WI) using standard protocols. Sequences of the inserts were determined in both directions using chain-terminating inhibitors and T7 DNA polymerase (T7 sequencing kit; Pharmacia Biotech Europe, Brussels) or by means of an ABI PRISM DNA sequencer 377 (Perkin-Elmer, Foster City, CA). In some cases, an additional terminal transferase extension was included to eliminate nonspecific stops appearing in the four lanes. The more internal sequences that could not be reached using vector-specific primers were determined by using primers complementary to internal regions of the CTV cDNA (Table 2).

Sequence analysis. CTV RNA sequences were compared using the programs GAP and BESTFIT (gap weight 5, length weight 0.3), multiple sequence alignments were generated by the PILEUP program, and a table of the pairwise distances was made with the

TABLE 1. Origin and pathogenic characteristics of the citrus tristeza virus isolates studied

Isolate	Geographical origin	ML[a]		CM[a]		Gpft[a]	SwO[a]		Decline SwO/SO[a]	
		VC[b]	SP[b]	VC	SP	SP	SP	SY[b]	Greenhouse[c]	Field
T308	Burjasot (Valencia)	3[d]	3	3	3	1	0	1	0	ND
T388	Japan	3	3	3	3	3	3	3	3	ND
T385	Orihuela (Alicante)	1	0	0	1	0	0	0	0	0
T317	Idem	1	1	2	2	0	0	0	0	ND
T318	Idem	3	3	3	3	ND	3	3	3	ND
T411	Idem	1	0	ND	ND	ND	0	ND	ND	ND
T408	Alzira (Valencia)	2	2	ND	ND	ND	0	0	ND	ND
T373	Alhama (Murcia)	2	1	2	1	0	0	0	0	2
T346	Brenes (Sevilla)	2	2	2	1	0	0	0	0	3

[a] Indicator species: ML (Mexican lime), CM (*Citrus macrophylla*), Gpft (Duncan grapefruit), SwO (Pineapple sweet orange), and SwO/SO (sweet orange grafted on sour orange).
[b] VC = vein-clearing, SP = stem-pitting, and SY = seedling yellows reaction on sour orange or Eureka lemon seedlings.
[c] Symptoms observed in plants grown in pots for less than 2 years in a temperature-controlled (18 to 26°C) greenhouse.
[d] Symptom intensity estimated in a 0 to 3 scale in which 0 = no symptoms and 3 = intense symptoms. ND = not determined.

TABLE 2. Primers used for cDNA synthesis, polymerase chain reaction amplification, and DNA sequencing

Name	Polarity	Primer sequence (5′ to 3′)[a]	Binding site with respect to T36 sequence[b]
PM1		ccggatcctctagagcggccgctttttttttttttttV[c]	
RF-92	–	CCGTAGAGGGACTATCGGC	266–284
RF-81	+	GTCGAAACTCAGAGGAAGC	110–128
RF-72	+	GTCTCTCCATCTTGCGTGTAG	18,292–18,312
RF-108	–	CCAGTCTCGTCTTCTCCCTTTCAGCG	18,813–18,838
RF-76	–	CAATCAGATGAAGTGGTG	19,009–19,026

[a] Virus-specific sequences are presented in uppercase letters.
[b] For the negative-strand primers, the coordinates are indicated for the positive strand.
[c] V = A, C, or G.

HOMOLOGIES program, all from the GCG sequence analysis package (Genetics Computer Group, Inc., Madison, WI) (7). A CLUSTAL W multiple sequence alignment (30) was used to construct an unrooted phylogenetic tree using the Fitch distance matrix program from the PHYLIP package (J. Felsenstein, University of Washington, Seattle). RNA secondary structures of lowest free energy were predicted using MFOLD (34).

RESULTS

Examination of the 5′-terminal region of CTV RNA: Isolates are complex mixtures of variants on the basis of sequence polymorphism. Comparison of the consensus CTV RNA sequences from isolates T36 (14) and VT (19) revealed that the nucleotide identity between their corresponding 5′ UTR of 107 and 105 nt, respectively, is only about 60% (19). To assess if this extreme polymorphism is a common feature among CTV isolates, the 5′-terminal regions from isolates T388, T317, T318, T411, and T408 were cloned and a total of 36 clones were sequenced, with each isolate being represented by at least five clones.

The length of the 5′ UTR of the 36 clones was 107 to 108 nt, except in some cases in which deletions at the 5′ end were observed (discussed below). Eight representative sequences from the different CTV isolates were aligned with the consensus sequences from the T36 and VT isolates (Fig. 1), and the levels of pairwise nucleotide identities of the 5′ UTRs were calculated (Table 3). The 36 clones clustered in three groups termed I, II, and III. The intragroup sequence identity was greater than 96% in group I, and greater than 88% in groups II and III. The intergroup sequence identity was 55 to 57% between groups I and II, 62 to 64% between groups II and III, and 44 to 45% between groups I and III (Table 3).

The T36 consensus sequence and four clones derived from isolate T388 belonged to group I, and the VT consensus sequence together with four and three clones derived from T388 and T318, respectively, belonged to group II. The remaining 25 clones, obtained from isolates T408 (six clones), T317 (ten clones), T411 (five clones), T318 (three clones), and T388 (one clone), belonged to group III. Additional sequences from isolate T385 (three clones) and from isolates T308, T373, and T346 (one clone each) were found to belong to groups III and I, respectively (data not shown). When the relationships among the 38 different sequences of the 5′ UTR including the T36 and VT consensus sequences were subjected to a phylogenetic reconstruction, the topology of the unrooted tree obtained (data not shown) was consistent with the clustering presented in Figure 1 and Table 3, supporting the robustness of the grouping.

A remarkable property of the 5′ UTR of the three groups of clones is the high content of A (27 to 35%) and C (28 to 34%) in combination with a low content of G (14 to 18%). Moreover, the multiple alignment of 5′-UTR sequences of CTV RNA revealed the presence of at least four CAA motifs (Fig. 1). The first CAA motif (positions 8 to 10 of the T36 consensus sequence) is strictly conserved in all clones, whereas the distribution of the remaining CAA repeats depends on the group to which the clone belongs.

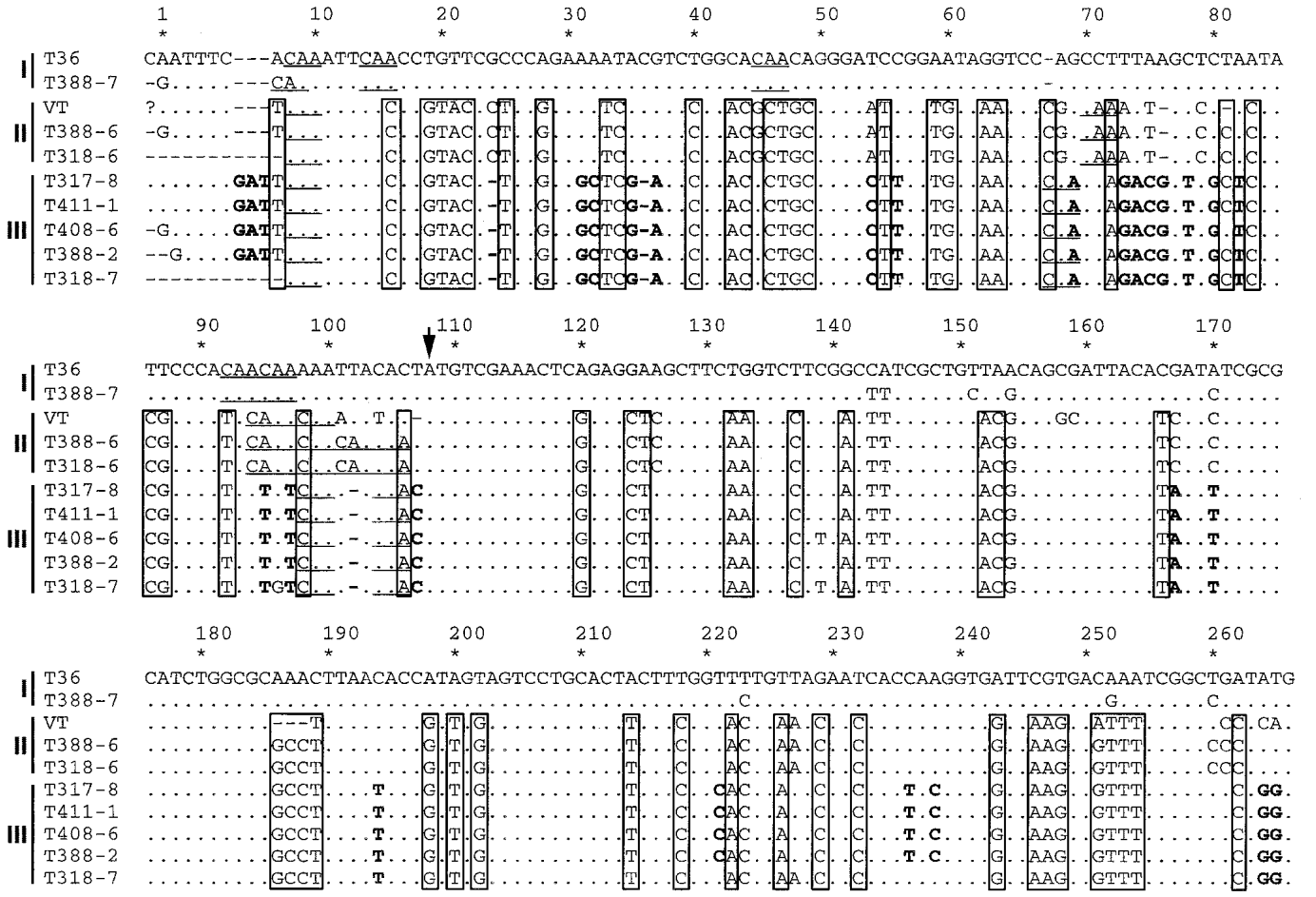

Fig. 1. Nucleotide sequence alignment with the program PILEUP of the 5′ untranslated region (UTR) and first 159 nucleotides of the open reading frame (ORF) 1a of eight representative clones of citrus tristeza virus (CTV) RNA from five different isolates. The two other available sequences of CTV RNA from T36 and VT isolates (14,19) are also included for comparative purposes. Dots indicate residues identical to the T36 sequence and dashes denote gaps. CAA motifs in the 5′ UTR are underlined and the ATG initiation codon of ORF 1a is indicated by an arrow. Boxed nucleotides are common to sequences of groups II and III, and those in bold are specific to sequences of group III. VT and T318-7 sequences are included within groups II and III, respectively, although they lack minor portions of the sequences defining these groups.

Multiple sequence alignments of the 5′-most 159 nt of ORF 1a, using nucleotide and amino acid sequences (Fig. 1 and data not shown, respectively), resulted in the same clustering as that previously obtained with the adjacent 5′-UTR sequences. The nucleotide sequence identity was greater than 95, 94, and 97% within sequences of groups I, II, and III, respectively. The intergroup sequence identity was 72% between groups I and III, 73 to 76% between groups I and II, and 91% between groups II and III. On the other hand, the amino acid sequence similarity was 67 to 72% between groups I and II, 70 to 71% between groups I and III, and 90 to 94% between groups II and III.

Heterogeneity of the 5′ end of CTV RNA. Previous results have shown that the 5′-terminal nucleotide of the genomic CTV RNA of the consensus sequences of isolates T36 and VT is an A (14,19), but when the sequence of the same region was determined in T36 after polyadenylation of the 3′ terminus of the dsRNA minus strand, an additional G was observed (14), shown as a C not numbered in the 5′-UTR sequence of the plus strand (Fig. 1). The presence of an unpaired G at the 3′ end of the minus strand of the dsRNA forms is a feature shared by several members of the alphavirus-like superfamily to which CTV belongs (33). However, only 21 out of the 36 clones reported here, also obtained by PCR amplification from the polydenylated minus strands of the dsRNAs, exhibited the sequence CAATTT (plus strand, in which the first A is position 1), suggesting that the unpaired G at the 3′ terminus of the minus strand in the CTV dsRNA is a frequent but not general characteristic. Other 5′ ends of the plus strand were GATTT, GTTT, AATTT, and CATTTT. A deletion of the seven 5′-terminal nucleotides of the T36 consensus sequence was observed in clones T318-6 and T318-7 of groups II and III, respectively (Fig. 1). Although it is possible that this deletion could result either from the degradation of the 5′ end of the template RNA or from the premature termination of its cDNA, two observations suggest that this may not be the case. First, the 5′ end of the 2.4-kilobase (kb) CTV-defective RNA characterized from the VT isolate contains exactly the same deletion (18), and second, the 5′ terminus resulting from the deletion was CAAATTC, very similar to the CAATTTC 5′ terminus observed in most CTV RNA sequences (Fig. 1).

Proposed secondary structure for the 5′ UTR of CTV RNA. Figure 2 shows the predicted secondary structures of lowest free energy for the 5′-UTR sequences. In spite of the low intergroup nucleotide identity, these structures were very similar and contained, in all cases, two stem-loop structures termed A and B. Stem-loop A has two helices of 6 base pairs (bp) separated by a C-rich internal loop of two out of three Cs in group I, five out of six Cs in group II, and five out of five Cs in group III. It should be noted that the 12 bp involved in the structures of groups II and III are identical, including a non-Watson-Crick G-U pair, while the structure of group I has five different base pairs, presumably resulting from compensatory mutations (Fig. 2). The sequences forming the terminal loop are variable, but of very similar sizes (6 or 7 nt).

Stem-loop B has a similar structure, with two helices of 6 and 3 to 5 bp separated by an internal loop larger than that of stem-loop A, but also rich in C: three out of nine, six out of ten (five out of nine in VT), and four out of eight Cs in structures of group I, II, and III, respectively. The helix of 6 bp is identical in groups II and III but not in group I, which has two different base pairs as a result of compensatory mutations (Fig. 2). The other helix has 3, 4, and 5 bp in groups I, II, and III, respectively; only 2 bp are identical in groups I and II, whereas the 5 bp in group III are different. The corresponding terminal loops are variable in sequence and length (3 to 6 nt).

When the alignment of the 5′-UTR sequences was modified to assume that positions structurally homologous in stem-loops A and B were also evolutionarily homologous, the allocation of the sequences to the three groups did not change (data not shown).

Examination of the 3′-terminal region of CTV RNA: Isolates comprise variants with the UTR sequence extremely conserved. The 3′ UTR of the CTV RNA extends along 272 and 273 residues in the VT and T36 consensus sequences, respectively (19,26), with a nucleotide identity of 97% between them. When these sequences were compared with the same region of the 15 clones from the CTV isolates T388, T317, T318, T411, and T408, with each isolate being represented by three clones, the nucleotide substitutions found were very limited and mainly located in the first third of this region, with the nucleotide identity always higher than 97% (data not shown). There is a minor discrepancy in the exact 3′ terminus of the genomic single-stranded RNA (ssRNA) reported for the T36 consensus sequence, which was proposed initially to be CCA while the plus strand of the dsRNA contained an extra U (14), and more recently to be CC with an extra A in the plus strand of the dsRNA form (15). Only 1 out of 15 clones here obtained by PCR amplification from the polyadenylated plus strands of the CTV dsRNAs had the sequence CCAT, while 11 clones exhibited the CCA sequence. The 3′ ends of the other three clones were CCAG, CTA, and CCCGA, suggesting that some sequence heterogeneity also exists in the 3′ terminus.

Conserved amino acid motifs in the potential p23 protein encoded by ORF 11. When the ORF 11 sequences (630 nt) of the 15 clones from the CTV isolates T388, T317, T318, T411, and T408 were compared with the homologous VT and T36 consensus sequences (19,26) and with 10 clones of this same region reported recently (25), a high degree of nucleotide identity (89 to 100%) was observed among all of them (data not shown). The alignment of amino acid sequences showed levels of similarity and identity higher than 92 and 85%, respectively. In all cases, the putative protein starts at similar positions (nt 18,394 of the T36 consensus sequence) and ends also at similar positions (nt 19,021 to 19,023

TABLE 3. Percentage of nucleotide sequence identities between the 5′ untranslated region of 10 representative citrus tristeza virus RNA sequences[a]

	I		II			III				
	T36	T388-7	VT	T388-6	T318-6	T317-8	T411-1	T318-7	T408-6	T388-2
I										
T36	…									
T388-7	96.3	…								
II										
VT	56.9	55.6	…							
T388-6	56.9	57.4	93.5	…						
T318-6	52.3	51.9	87.9	93.5	…					
III										
T317-8	44.7	43.8	63.1	62.5	61.6	…				
T411-1	45.0	44.2	63.6	63.1	62.2	99.1	…			
T318-7	45.0	44.5	63.6	63.0	67.3	89.0	89.8	…		
T408-6	45.0	44.2	63.6	63.1	62.2	98.2	99.1	89.8	…	
T388-2	45.5	44.6	64.2	63.6	62.8	97.3	98.2	90.7	98.2	…

[a] The values were estimated with the program HOMOLOGIES including gaps in the comparisons. Sequences for T36 and VT have been reported previously (14 and 19, respectively).

of the T36 consensus sequence), with a constant size of 209 amino acids approximately equivalent to 23 kDa (p23).

Although the deduced amino acid sequence did not show any significant similarity to other sequences in databases (26), the presence of a cluster of positively charged amino acid residues and a pair of conserved cysteines, also observed in proteins with RNA-binding properties encoded at the 3′ end of the genome of carlavirus and several related viruses (13,17), suggested that p23 could also be an RNA-binding protein involved in the regulation of virus gene expression (8). A more detailed examination has led us to observe that all the available p23 sequences have the motif CVDCGRKHDKALKTERKC between amino acids 68 and 85. The underlined cysteine and histidine residues and their relative positions within the motif make them very good candidates for forming a tetrahedral "zinc-finger" domain (20). Moreover, the analysis of the hydropathy profile and the distribution of the acidic and basic amino acids has revealed that the N-terminal domain of p23 has a net positive charge (the pI of the region encompassing amino acids 29 to 155 is 10.99), whereas the C-terminal domain has a net negative charge (the pI of the region covering the last 54 amino acids is 4.35).

The sequences of 81 to 82 residues of the intercistronic region (IR) between ORF 10 and ORF 11 obtained for the same 15 clones indicated above were compared with the consensus sequences of the VT and T36 isolates (19,26). The nucleotide identity was always higher than 90%. A remarkable property of this IR is its high content (68%) of A+T residues.

DISCUSSION

The analysis of the variability observed in the 5′ UTR allowed a classification of the sequences into three groups (I, II, and III) represented by the sequences of T36, VT, and clone T317-8, respectively (Fig. 1). While the intragroup nucleotide identity was higher than 88%, the intergroup nucleotide identity ranged between 44 and 63%, setting limits for the assignment of each of these sequences to one of the three groups. Moreover, these data indicate that the spectrum of sequence variation of the 5′ UTR is not continuous, but concentrated in peaks separated by valleys, a general feature of viral quasispecies (10). The analysis of sequence polymorphism of the 5′ region of ORF 1a led also to the same groups I, II, and III, although the intra- and intergroup sequence identities were higher than those found between the corresponding 5′ UTRs, a clear reflection that both types of sequences are under different functional constraints.

It is interesting to note that some of the CTV isolates contained sequences belonging to more than one group, in the case of T388 from Japan, to all three groups. Under field conditions, successive infections by different CTV inocula could explain the presence of sequences from more than one group in the same isolate. It should also be emphasized that the actual polymorphism could be even higher, because the RT-PCR approach used to obtain the clones may favor the amplification of specific sequences. Although only one sequence has been reported for isolates T36 and VT (14,19), as well as for T308, T373, and T346 (this work), it is likely that some of them may contain 5′-terminal sequences belonging to other groups. Also worthy of note from the point of view of the geographical origin is the observation that most of the sequences present in Spanish CTV isolates belong to group III, in contrast to isolate T388 from Japan, which is composed mostly of sequences of groups I and II. From a phytopathological standpoint, it is intriguing that the mildest isolates studied here (T385, T317, T411, and T408) contained only sequences of group III, whereas the

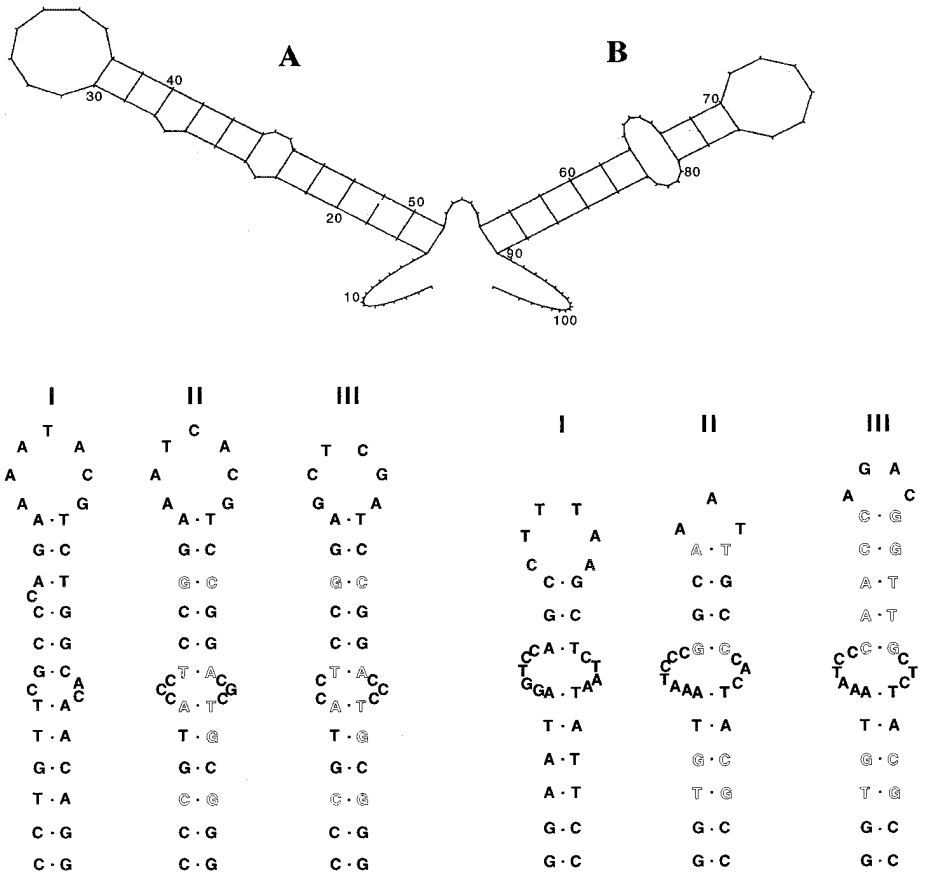

Fig. 2. Predicted secondary structures of lowest free energy of the 5′ untranslated region of citrus tristeza virus RNA using the MFOLD program. Top, Schematic representation of the secondary structure of the T36 sequence. Bottom, Primary and secondary structures of stem-loops A (left) and B (right) of representative sequences of groups I (T36), II (VT), and III (T317-8). Outlined fonts indicate single- or double-compensatory mutations with respect to the T36 sequence that preserve the structure of stem-loops A and B.

most severe isolates (T318 and T388) also contained sequences of groups I, II, or both. However, specific symptoms cannot be attributed to any of the three groups of sequences at this time.

The genomic RNA of closteroviruses functions as a messenger and as a template RNA, but the specific sequences involved in the regulation of these functions are unknown. However, a detailed inspection of the 5′ UTR of the CTV RNA has revealed conserved CAA motifs identified previously in the 5′ UTR of other viral RNAs and shown to be functionally significant (12,16,31). Whether the CAA motifs by themselves, or in combination with other elements, may play some functional role in the case of the CTV RNA remains to be determined.

The study of the secondary structure of the 5′ UTR of CTV RNA by a combination of sequence comparison and computer predictions has led to a general model with two stem-loops (Fig. 2). Interestingly, most of the variability is accommodated in the loops and, when found in the stems, occurred in such a way that compensatory mutations maintained their general features. This conservation of the secondary structure by covariation strongly supports a functional in vivo role for it. Hairpin structures provide potential sites for nucleation in RNA folding and for interactions with proteins (32).

In contrast to the 5′-terminal region, the variability found in the last 3′-terminal 984 nt of the CTV RNA was very much restricted, with nucleotide identity values of 90% or higher irrespective of the geographical origin or biological properties of the isolates. Therefore, the very different levels of sequence conservation observed between the 5′- and 3′-terminal regions of the T36 and VT sequences (19) may be typical of sequences from other CTV isolates as well. The presence of a putative "zinc-finger" domain in p23 and other structural properties predicted for this protein are typical of the transactivating domains of many transcription factors (6). It is tempting to speculate that p23, which has been recently found in the cytoplasmic fraction of CTV-infected tissue (25), might play a regulatory role in the replication of CTV RNA. Circumstantial evidence supporting this speculation is the early accumulation of the p23 subgenomic RNA in infected protoplasts (24).

Genomic determinants responsible for the different pathogenic features of CTV are currently unknown. Suggested correlations between amino acid sequences of the p23 and coat proteins (25, 27) have been based on single sequences from each isolate. However, our results show that sequence variants, in some cases distantly related, coexist within the same isolate, making it difficult to assess the validity of such correlations. Unambiguous assignment of biological traits to specific viral genomic regions awaits the availability of infectious full-length cDNA clones.

Although sequencing of CTV RNA from new isolates might increase the complexity of the current pattern, the classification of sequences proposed here on the basis of the 5′-terminal region of the genomic RNA provides the first comprehensive framework for accommodating the extreme variability observed in this viral RNA.

ACKNOWLEDGMENTS

This work was partially supported by grants from the Conselleria de Agricultura y Medio Ambiente, Conselleria de Cultura, Educación y Ciencia (GV-3116/95), INIA (project SC-93111), as well as by contract ERBIC18CT960044 from the European Union. M. A. Ayllón was the recipient of a fellowship from the Ministerio de Educación y Ciencia. We thank W. O. Dawson, V. Pallás, and C. Hernández for the critical reading of the manuscript and for valuable suggestions; and M. E. Martínez and M. Boix for excellent technical assistance.

LITERATURE CITED

1. Agranovsky, A., Lesemann, D. E., Maiss, E., Hull, R., and Atabekov, J. G. 1995. "Rattlesnake" structure of a filamentous plant RNA virus built of two capsid proteins. Proc. Natl. Acad. Sci. U.S.A. 92:2470-2473.
2. Ballester-Olmos, J. F., Pina, J. A., Carbonell, E., Moreno, P., Hermoso de Mendoza, A., Cambra, M., and Navarro, L. 1993. Biological diversity of citrus tristeza virus (CTV) isolates in Spain. Plant Pathol. 42:219-229.
3. Ballester-Olmos, J. F., Pina, J. A., and Navarro, L. 1988. Detection of a tristeza-seedling yellows strain in Spain. Pages 28-32 in: Proc. Conf. IOCV, 10th. International Organization of Citrus Virologists, Riverside, CA.
4. Bar-Joseph, M., and Lee, R. F. 1989. Citrus tristeza virus. CMI/AAB Descriptions Plant Viruses 353.
5. Bar-Joseph, M., and Loebenstein, G. 1973. Effect of strain, source plant and temperature on the transmissibility of citrus tristeza virus by the melonaphid. Phytopathology 63:716-720.
6. Coleman, J. E. 1992. Zinc proteins: Enzymes, storage proteins, transcription factors and replication proteins. Annu. Rev. Biochem. 61:897-946.
7. Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.
8. Dolja, V. V., Karasev, A. V., and Koonin, E. V. 1994. Molecular biology and evolution of closteroviruses: Sophisticated build-up of large RNA genomes. Annu. Rev. Phytopathol. 32:261-285.
9. Domingo, E., Escarmís, C., Sevilla, N., Moya, A., Elena, S. F., Quer, J., Novella, I. S., and Holland, J. J. 1996. Basic concepts in RNA virus evolution. FASEB (Fed. Am. Soc. Exp. Biol.) J. 10:859-864.
10. Eigen, M., and Biebricher, C. K. 1988. Sequence space and quasispecies distribution. Pages 211-245 in: RNA Genetics, Vol. 3. E. Domingo, J. J. Holland, and P. Alquist, eds. CRC Press, Boca Raton, FL.
11. Febres, V. J., Ashoulin, L., Mawassi, M., Frank, A., Bar-Joseph, M., Manjunath, K. L., Lee, R. F., and Niblett, C. L. 1996. The p27 protein is present at one end of citrus tristeza virus particles. Phytopathology 86: 1331-1335.
12. Gallie, D. R., and Walbot, V. 1992. Identification of the motifs within the tobacco mosaic virus 5′-leader responsible for enhancing translation. Nucleic Acids Res. 20:4631-4638.
13. Gramstat, A., Courtpozanis, A., and Rohde, W. 1990. The 12 kDa protein of potato virus M displays properties of a nucleic acid-binding regulatory protein. FEBS (Fed. Eur. Biochem. Soc.) Lett. 276:34-38.
14. Karasev, A. V., Boyko, V. P., Gowda, S., Nikolaeva, O. V., Hilf, M. E., Koonin, E. V., Niblett, C. L., Cline, K., Gumpf, D. J., Lee, R. F., Garnsey, S. M., Lewandowski, D. J., and Dawson, W. O. 1995. Complete sequence of the citrus tristeza virus RNA genome. Virology 208:511-520.
15. Karasev, A. V., Nikolaeva, O. V., Mushegian, A. R., Lee, R. F., and Dawson, W. O. 1996. Organization of the 3′-terminal half of beet yellow stunt virus genome and implications for the evolution of closteroviruses. Virology 221:199-207.
16. Kim, K.-H., and Hemenway, C. 1996. The 5′ nontranslated region of potato virus X RNA affects both genomic and subgenomic RNA synthesis. J. Virol. 70:5533-5540.
17. Koonin, E. W., Boyko, V. P., and Dolja, V. V. 1991. Small cysteine-rich proteins of different groups of plant RNA viruses are related to different families of nucleic acid-binding proteins. Virology 181:396-398.
18. Mawassi, M., Karasev, A. V., Mietkiewska, E., Gafny, R., Lee, R. F., Dawson, W. D., and Bar-Joseph, M. 1995. Defective RNA molecules associated with citrus tristeza virus. Virology 208:383-387.
19. Mawassi, M., Mietkiewska, E., Gofman, R., Yang, G., and Bar-Joseph, M. 1996. Unusual sequence relationships between two isolates of citrus tristeza virus. J. Gen. Virol. 77:2359-2364.
20. Miller, J., McLachlan, A., and Klug, A. 1985. Repetitive zinc-binding domains in the protein transcription factor III A from *Xenopus oocytes.* EMBO (Eur. Mol. Biol. Organ.) J. 4:1609-1614.
21. Moreno, P., Guerri, J., Ballester-Olmos, J. F., Albiach, R., and Martínez, M. E. 1993. Separation and interference of strains from a citrus tristeza virus isolate evidenced by biological activity and double-stranded RNA (dsRNA) analysis. Plant Pathol. 42:35-41.
22. Moreno, P., Guerri, J., Ballester-Olmos, J. F., Fuertes-Polo, C., Albiach, R., and Martínez, M. E. 1993. Variations in pathogenicity and double-stranded RNA (dsRNA) patterns of citrus tristeza virus isolates induced by host passage. Pages 8-15 in: Proc. Conf. IOCV, 12th. International Organization of Citrus Virologists, Riverside, CA.
23. Moreno, P., Guerri, J., and Muñoz, N. 1990. Identification of Spanish strains of citrus tristeza virus (CTV) by analysis of double-stranded RNAs (dsRNA). Phytopathology 80:477-482.
24. Navas-Castillo, J., Albiach-Martí, M. R., Gowda, S., Hilf, M. E., Garnsey, S. M., and Dawson, W. O. 1996. Kinetics of accumulation of citrus tristeza virus RNAs. Virology 228:92-97.
25. Pappu, S. S., Febres, V. J., Pappu, H. R., Lee, R. F., and Niblett, C. L. 1997. Characterization of the 3′ proximal gene of the citrus tristeza closterovirus genome. Virus Res. 47:51-57.
26. Pappu, H. R., Karasev, A. V., Anderson, E. J., Pappu, S. S., Hilf, M. E., Febres, V. J., Eckloff, R. M. G., McCaffery, M., Boyko, V., Gowda, S., Dolja, V. V., Koonin, E. V., Gumpf, D. J., Cline, K. C., Garnsey, S. M., Dawson, W. O., Lee, R. F., and Niblett, C. L. 1994. Nucleotide sequence and organization of eight 3′ open reading frames of the citrus tristeza

closterovirus genome. Virology 199:35-46.
27. Pappu, H. R., Pappu, S. S., Niblett, C. L., Lee, R. F., and Civerolo E. 1997. Comparative sequence analysis of the coat proteins of biologically distinct citrus tristeza closterovirus isolates. Virus Genes 7:255-264.
28. Rosner, A., Lee, R. F., and Bar-Joseph, M. 1986. Differential hybridization with cloned cDNA sequences for detecting a specific isolate of citrus tristeza virus. Phytopathology 76:820-824.
29. Sekiya, M. E., Lawrence, S. D., McCaffery, M., and Cline, K. 1991. Molecular cloning and nucleotide sequencing of the coat protein gene of citrus tristeza virus. J. Gen. Virol. 72:1013-1020.
30. Thompson, J. D., Higgins, D. G., and Gibbson, T. J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.
31. Tordo, V. M. J., Chachulska, A. M., Fakhfakh, H., Le Romancer, M., Robaglia, C., and Astier-Manifacier, S. 1995. Sequence polymorphism in the 5′ NTR and in the P1 coding region of potato virus Y genomic RNA. J. Gen. Virol. 76:939-949.
32. Witherell, G. W., and Uhlenbeck, O. C. 1989. Specific RNA binding of Qβ coat protein. Biochemistry 28:71-76.
33. Wu, G., and Kaper, J. M. 1994. Requirement of 3′-terminal guanosine in (–)-stranded RNA for in vitro replication of cucumber mosaic virus satellite RNA by viral RNA-dependent RNA polymerase. J. Mol. Biol. 238:655-657.
34. Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecule. Science 224:48-52.